

SUPPLEMENTARY OF “POSTERIOR-GUIDED VISUAL TOKEN PRUNING IN VISION–LANGUAGE MODELS”

Anonymous authors

Paper under double-blind review

1 EXPERIMENTS

1.1 QUALITATIVE RESULTS

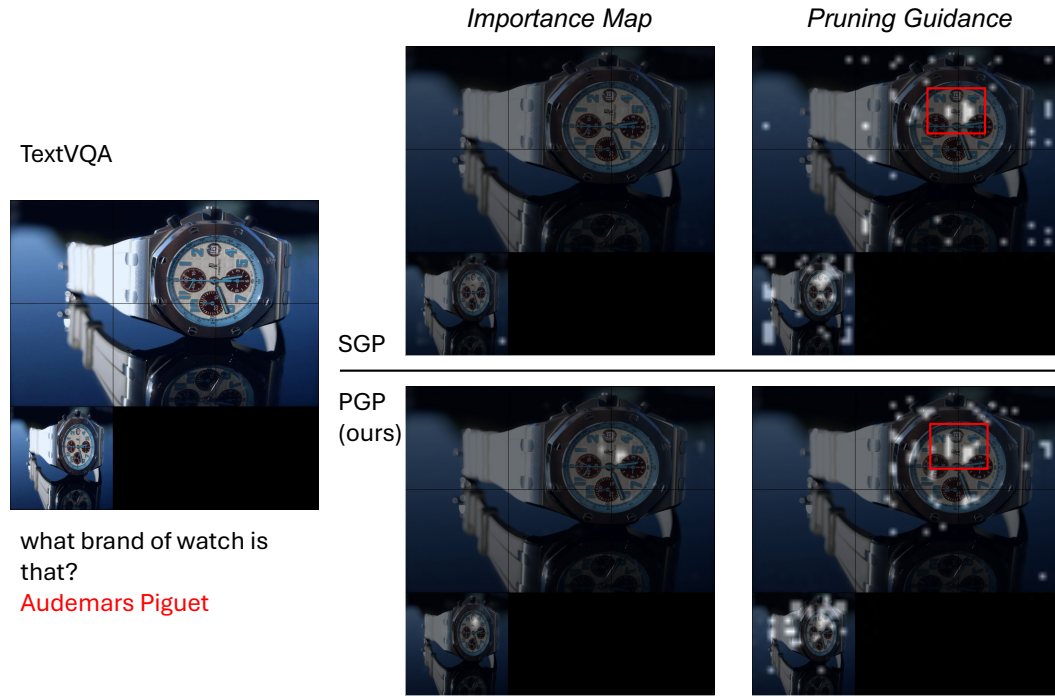
In Fig. 1, Fig. 2, and Fig. 3, we provide a comparison of the importance map and the pruning guidance provided by SGP and our proposed PGP. PGP tends to assign a higher importance score to a wider and more relevant visual tokens than SGP. In Fig. 4 and Fig. 5, we present the qualitative results of PGP on the MMStar dataset, demonstrating the generalizability of PGP on real-world images.

1.2 HYPERPARAMETER

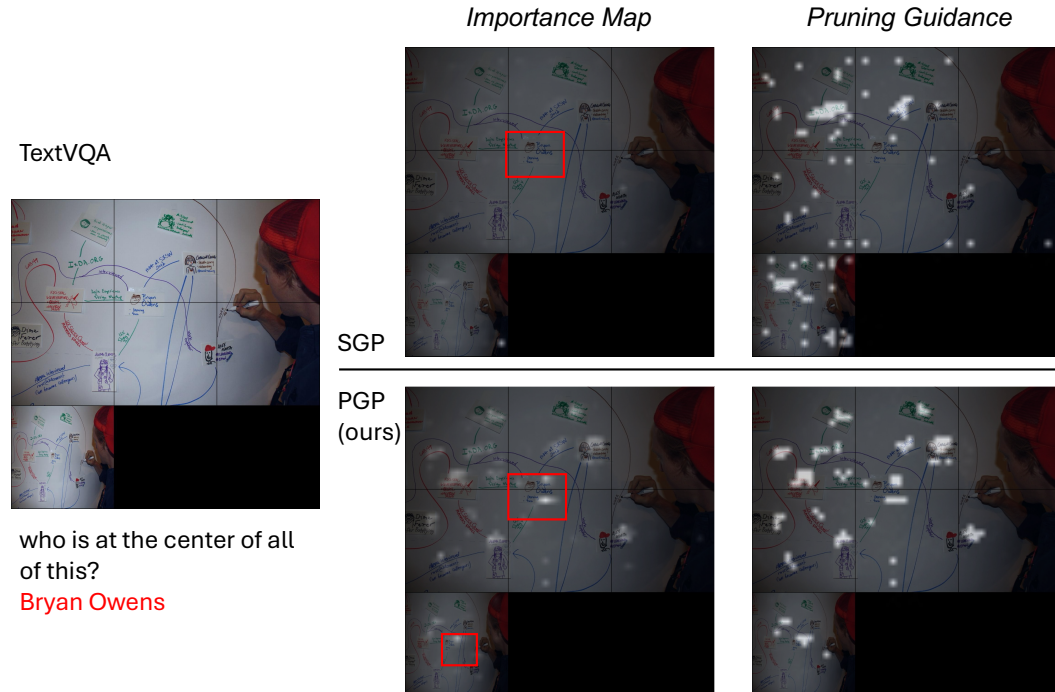
We detail the hyperparameters used for training the small model and the information bottleneck. To be noticed, the large model was not involved during training. The only trainable parts include the LoRA and the information bottleneck module.

Table T1: Hyperparameters for training.

LoRA alpha	64
LoRA rank	32
Batch size (SFT)	16
Gradient accumulation (SFT)	11
Learning rate	0.00005
Optimizer	AdamW
Weight decay	0.01
γ	$0.2 * total\ steps$
τ_{max}	0.5
τ_{min}	0.2



078 Figure 1: Comparison of the importance map and pruning guidance proposed by SGP and PGP
079 (ours) based on the user input.



105 Figure 2: Comparison of the importance map and pruning guidance proposed by SGP and PGP
106 (ours) based on the user input.

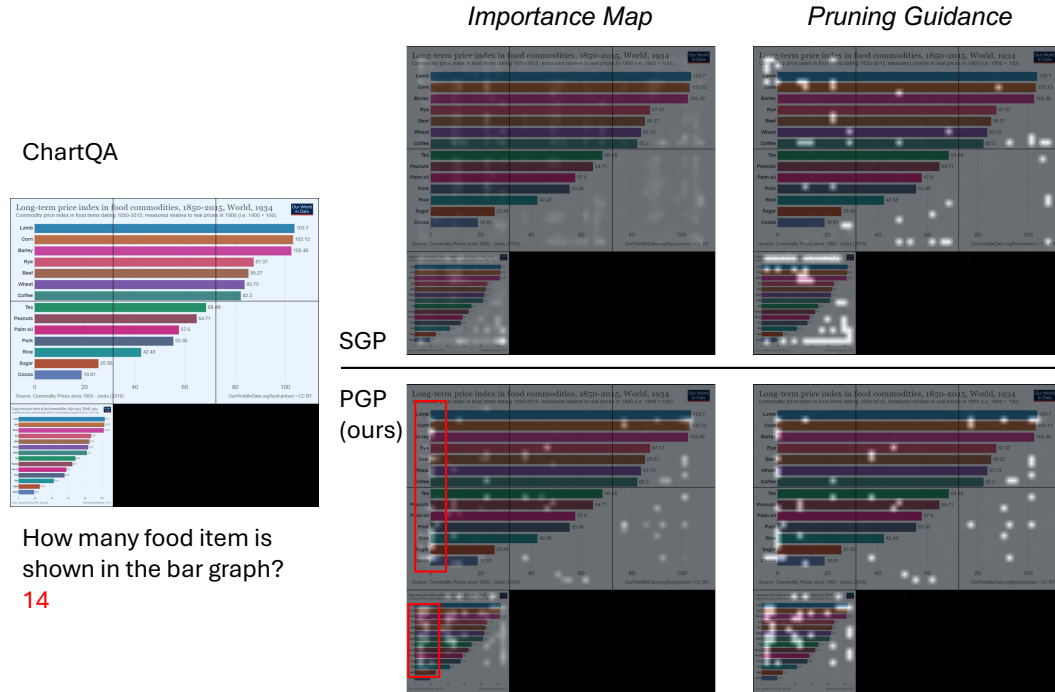


Figure 3: Comparison of the importance map and pruning guidance proposed by SGP and PGP (ours) based on the user input.

Which option describe the object relationship in the image correctly.

- A: The suitcase is on the book.,
 B: The suitcase is beneath the cat.,
 C: The suitcase is beneath the bed.,
 D: The suitcase is beneath the book.

What is the sport being played in the image?

- A: Tennis,
 B: Soccer,
 C: Volleyball,
 D: Basketball

Answer with the option's letter from the given choices directly

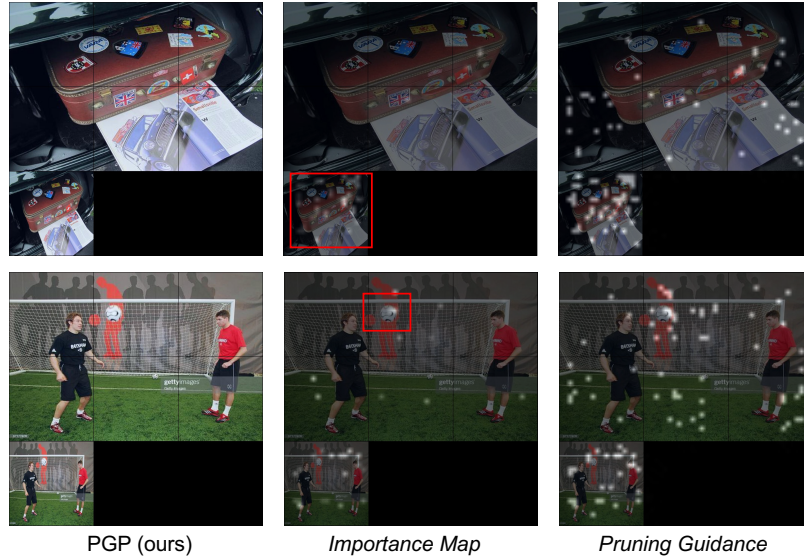


Figure 4: Visualization of the importance map and pruning guidance proposed by PGP (ours) based on the user input.

Which action is
performed in this
image?
Long jump

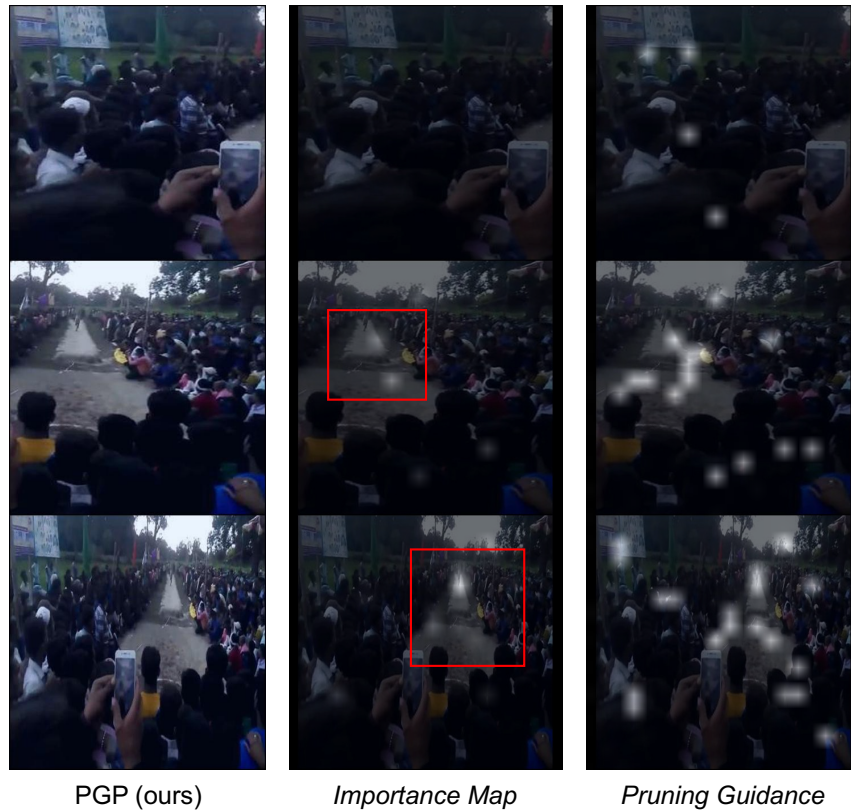


Figure 5: Visualization of the importance map and pruning guidance proposed by PGP (ours) based on the user input.